

## YUSHI TANG

E-mail: ytang@hsph.harvard.edu  
Phone: 857-209-1273

Harvard T.H. Chan School of Public Health  
Department of Biostatistics  
655 Huntington Avenue  
Boston, MA 02115

### Education

From 2019	<i>Princeton University</i> <b>Ph.D. Candidates in Quantitative and Computational Biology</b>	Princeton, NJ, USA
2017 - 2019	<i>Harvard T. H. Chan School of Public Health (HSPH)</i> <b>M.S. in Computational Biology and Quantitative Genetics</b>	Boston, MA, USA
2016 - 2017	<i>Department of Statistics, UCLA</i> <b>Cross-disciplinary Scholars in Science and Technology</b>	Los Angeles, CA, USA
2013 - 2017	<i>Peking University (PKU)</i> <b>B.S. in Environment Sciences</b> Concentration in Environmental Statistics <b>B.A. in Economics</b>	Beijing, P.R.China

### Research Experience

From 2018	<b>RA, Department of Biostatistics, HSPH</b> <b>The Genome Sequencing Program of the National Human Genome Research Institute</b> <ul style="list-style-type: none"> <li>◆ Developed methods and software for the gold-standard quality control (QC) procedure to analyze large-scale whole-genome sequencing data.</li> <li>◆ Designed protocols and algorithms for predicting sample population, identifying duplicated samples, and making accurate sex inference based on GSP's genotype data.</li> </ul>	Supervisor: Prof. Xihong Lin
2016 - 2017	<b>CSST Scholar &amp; RA, Department of Statistics, UCLA</b> <b>ENCODE-DREAM Challenge: Predictions on <i>in vivo</i> Transcription Factor Binding Site</b> <ul style="list-style-type: none"> <li>◆ Identified 4 crucial features from 14 original variables in ENCODE database for prediction including the accessibility of chromatin, the PWM scores of specific motifs, the DNA 3D Shape parameters, and k-mers of DNA sequence.</li> <li>◆ Developed a two-round selection model by using Bayesian Inference and Random Forests, which performed well for predictions on <i>in vivo</i> transcription factor binding site.</li> </ul>	Supervisor: Prof. Jingyi Jessica Li
2015 - 2017	<b>RA, Lab of Water Environmental Biotechnology, PKU</b> <b>Bio-information Analysis of Microbial Community in the Sediment of Hangzhou Bay</b> <ul style="list-style-type: none"> <li>◆ Developed a tripartite network and dimensional reduction method (TMEN) for integral analysis of 16S rRNA data, large-scale metagenomic data, and high-dimensional environmental data.</li> <li>◆ Applied TMEN to studying the marine microbial-environmental ecological network and discovered two ecological mechanisms including how the microbe become crucial species in a complex microbial-ecological network, as well as how to identify key taxonomic categories in complex microbial communities.</li> </ul>	Supervisor: Prof. Donghui Wen

### Awards and Honors

2019	<b>Lewis-Sigler Institute Scholars Award</b> in Quantitative and Computational Biology Awarded by Lewis-Sigler Institute, Princeton University
2017	<b>Outstanding Graduate of Beijing</b> as a Bachelor Awarded by the Beijing Municipal Commission of Education
2017	<b>Excellent Graduate</b> who has demonstrated outstanding performance Awarded by Peking University
2016 & 2015	<b>National Scholarship (2/1000)</b> Awarded by Ministry of Education of the People's Republic of China
2016 & 2015	<b>Merit Student Pacesetter</b>

- Awarded by Peking University
- 2016 **UCLA Cross-disciplinary Scholars in Science and Technology Award**  
In recognition of outstanding research and presentation skills
- 2014 - 2017 **Peking University Outstanding Researchers Program (Spark Program) Fellowship**  
Campus-wide honor awarded by PKU
- 2014 - 2017 **Top-notch Students Training in Basic Disciplines Fellowship in China**  
Nationwide honor awarded by the Ministry of Education of China

## Publications

- A1 **Tang Y**, Dai T, Su Z, Hasegawa K, Tian J, Chen L, and Wen D. (2018) Tripartite community structure of microbial-environment network indicates how crucial species influence the microbial community ecology [J]. *Microbial Ecology*. (Accept, 2019)
- A2 Zhao M, **Tang Y**, Kim H, and Hasegawa K. (2018) Machine learning with k-means dimensional reduction for predicting survival outcomes in patients with breast cancer [J]. *Cancer Informatics*. 2018 Nov. DOI: 10.1177/1176935118810215
- A3 Dai T, Zhang Y, Ning D, Su Z, **Tang Y**, Huang B, Mu Q, and Wen D. (2018) Dynamics of sediment microbial functional capacity and community interaction networks in an urbanized coastal estuary [J]. *Frontiers in Microbiology*. 2018 Nov. DOI: 10.3389/fmicb.2018.02731
- A4 Su Z, Dai T, **Tang Y**, Tao Y, Huang B, Mu Q, and Wen D. (2018) Sediment bacterial community structures and their predicted functions implied the impacts from natural processes and anthropogenic activities in coastal area [J]. *Marine Pollution Bulletin*. 2018 Apr. DOI: 10.1016/j.marpolbul.2018.04.052
- A5 Dai T, Zhang Y, **Tang Y**, Bai Y, Tao, Y, Huang B, and Wen D. (2016) Identifying the key taxonomic categories that characterize microbial community diversity using full-scale classification: a case study of microbial community in the sediments of Hangzhou Bay [J]. *FEMS Microbiology Ecology*. 2016 Oct. DOI: 10.1093/femsec/fiw150

## Conference Proceedings

- C1 Dai T, Zhang Y, **Tang Y**, and Wen D. (2016) Metagenomics reveals microbial community dynamics under natural and anthropogenic impact in the sediment of Hangzhou Bay, China [C]. The 3rd Xiamen Symposium on Marine Environmental Sciences (XMAS-III).  
[http://mel.xmu.edu.cn/conference/3XMAS/program\\_info.asp?id=251](http://mel.xmu.edu.cn/conference/3XMAS/program_info.asp?id=251)

## Teaching Experience

- 2019 Spring Department of Statistics, Harvard University Instructor: Dr. X. Shirley Liu  
*Teaching Assistant*: STAT 215 / BST 282 Intro to Computational Biology and Bioinformatics
- 2018 Summer Department of Biostatistics, HSPH Instructor: Dr. Garrett Fitzmaurice  
*Teaching Assistant*: BST 215 Linear and Longitudinal Regression

## Computer Skills

- Programming languages Proficient in C, Python, R, and MATLAB
- Cluster Computing Software Familiar with AWS, GCP, Odyssey, and Orchestra
- Basic statistics: SAS, Stata, Origin, and SPSS
- Bioinformatics: Bioconductor, MeV, LIMMA, DAVID, COMBAT, FastQC, Seurat, and DESeq
- Network Analysis: Cytoscape
- GWAs and Statistical Genetics: PLINK
- Others Familiar with databases including PUBMED, ENCODE, KEGG, UCSC, GEO, and cBioPortal

---

## Lab Skills

Molecular Biology	PCR, qPCR, site-directed mutagenesis
Chemistry	Frequent user of GC-MS, GC-MS/MS, ultraviolet spectrophotometer, fluorescence spectrophotometer and Abbe refractometer

---

## Organizational Experience

2016 - 2017	<b>The 34th Students' Union of Peking University</b> The undergraduate standing committee member
2015 - 2016	<b>The 8th Students' Union of College of Environmental Sciences and Engineering, PKU</b> The undergraduate vice-chair of academics
2015 - 2017	<b>Debating Team of College of Environmental Sciences and Engineering, PKU</b> Captain and coach
2015 - 2016	<b>The Accordion Band of PKU</b> Permanent conductor and the previous first accordion
2013 - 2017	<b>The Chorus of College of History &amp; CESE in Peking University</b> Chorus conductor

---

## Interests

Accordion, Chorus & Symphony Conductor, Debate, Swimming, Badminton